

Homework 2: Find three unusual points in NY property data

1. Zip code of city of New York should start with 1. However, there are three addresses start with 3, with the street name in New York Cit. One of them is owned by the city of New York.

OWNER	STADDR	ZIP
THE CITY OF NEW YORK	WEST 16 STREET	33803.0
NEPTUNE .	WEST 16 STREET	33803.0
NEPTUNE .	WEST 16 STREET	33803.0

2. Generally from the descriptive information, we understand the mean total value of tax class 1B ranked the third from the bottom. However, by looking at the mean exemption, we can find tax class 1B ranked number six, just below tax class 2C.

```
mydata[['TAXCLASS', 'AVTOT', 'EXTOT']].groupby(['TAXCLASS']).mean().sort_values(by='AVTOT', ascending=False)
```

	AVTOT	EXTOT
TAXCLASS		
4	1.492342e+06	760037.415434
1D	1.166866e+06	226899.206897
2	3.481780e+05	85687.720762
2B	1.772947e+05	27078.343575
2C	1.170227e+05	25635.414914
2A	7.944702e+04	6719.103999
1C	2.892319e+04	3314.697674
1	2.488275e+04	2965.347853
1B	1.448514e+04	8874.340242
1A	1.441980e+04	4672.965118
3	3.745119e+02	235.631324

To further understand it, we can see that among the list of tax class 2(apartment) and tax class 4(others), there's one property located at zip code 11359 (Fort Totten Park) that earned a lot of exemptions.

```
mydata[['TAXCLASS', 'ZIP', 'AVTOT', 'EXTOT']].groupby(['TAXCLASS', 'ZIP']).mean().sort_values(by='EXTOT', ascending=False)
```

	AVTOT	EXTOT
TAXCLASS ZIP		
10048.0	1.701000e+08	1.701000e+08
10281.0	1.180500e+08	1.180500e+08
4	10282.0	1.005094e+08
11422.0	2.596662e+07	2.573827e+07
2	10475.0	2.615507e+07
10044.0	2.033794e+07	1.953269e+07
4	11697.0	1.693358e+07
1B	11359.0	1.678021e+07
11005.0	1.467000e+08	1.556827e+07
2	10282.0	1.552380e+07
4	10028.0	1.462094e+07
2	11239.0	1.022487e+07
10280.0	6.902190e+06	6.764340e+06
4	11370.0	6.452418e+06
2	11434.0	5.968379e+06

3. Most of the properties without owner records that receive some tax exemptions are apartments (tax class 2) gathered around Brooklyn bridge (zip code 11211, 11201).

```
no_owner[['ZIP', 'EXLAND']].groupby(['ZIP']).count().sort_values(by='EXLAND', ascending=False).iloc[0:10,]
```

EXLAND	
ZIP	
11211.0	2047
11201.0	1536
10280.0	768
11215.0	704
10023.0	702

EXLAND		
TAXCLASS	ZIP	
2	11211.0	1431
	11201.0	1132
	10280.0	755
	10023.0	667
	11101.0	483